

Sabri A. Mahmoud,

"Arabic Character Recognition using Fourier Descriptors and Character Contour Encoding."

Pattern Recognition Journal, Vol. 27, No. 6, June 1994, pp. 815-824.

Key words: Arabic Character Recognition, OCR, Fourier Descriptors, Contour Analysis, Curvature Features, Direction Features



Pattern Recognition, Vol. 27, No. 6, pp. 815-824, 1994
Elsevier Science Ltd
Copyright © 1994 Pattern Recognition Society
Printed in Great Britain. All rights reserved
0031-3203/94 \$7.00 + .00

0031-3203(93)E0028-6

ARABIC CHARACTER RECOGNITION USING FOURIER DESCRIPTORS AND CHARACTER CONTOUR ENCODING

SABRI A. MAHMOUD

Computer Engineering Department, College of Computers and Information Sciences, King Saud University,
P.O. Box 51405, Riyadh 11543, Saudi Arabia

(Received 25 January 1993; in revised form 24 November 1993; received for publication 2 December 1993)

Abstract—Normalized Fourier descriptors are known to be invariant to scale, translation, and rotation. This technique was used by researchers of Latin OCR yielding acceptable results. In addition, contour analysis was used in object recognition with success. Both techniques are adopted as they are necessary for the recognition of Arabic characters with acceptable recognition rates. This combination was deemed necessary due to the special characteristics of Arabic characters that have some very similar characters. The character images are smoothed by a statistically-based algorithm to eliminate noise. Then, the contours of the image (namely the character primary part, the dots, and hole contours) are extracted. Fourier descriptors and curvature features of the primary part of the character are computed. These features of the training set are used as the model features. The features of an input character are compared to the models' features using a distance measure. The model with the minimum distance is taken as the class representing the character. The dots' and holes' features are then used to specify the particular character. Experimental results have shown that the combination of the Fourier descriptors, the curvature features and the use of dots' and holes' features to be powerful in successfully classifying Arabic characters. Recognition rates of 100% were achieved for the model classes. However, this rate has come down to 98% in the post-recognition phase of identifying the specific characters. The major part of these errors come from corrupted data.

Arabic character recognition OCR Fourier descriptors Contour analysis
Curvature features Direction features